ELSEVIER

Theoretical Computer Science 285 (2002) 101–117

Theoretical Computer Science

www.elsevier.com/locate/tcs

The convergence of stochastic algorithms solving flow shop scheduling ☆

K. Steinhöfel[a,b,*], A. Albrecht[c], C.K. Wong[d,1]

[a] *GMD, National Research Center for Information Technology, Kekuléstr. 7, D-12489 Berlin, Germany*

[b] *Institute for Theoretical Computer Science, Swiss Federal Institute of Technology Zurich, CH-8092 Zurich, Switzerland*

[c] *Department of Computer Science, University of Hertfordshire, Hatfield, Herts AL10 9AB, UK*

[d] *Department of Computer Science and Engineering, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong*

Received May 2001

Abstract

In the paper, we apply logarithmic cooling schedules of simulated annealing-based algorithms to flow shop scheduling. In our problem setting, the objective to minimize the overall completion time which is called the makespan. We prove a lower bound for the number of steps that are sufficient to approach an optimum solution with a certain probability. The result is related to the maximum escape depth Γ from local minima of the underlying energy landscape. In our approach, we need $n^{O(\Gamma)} + \log^{O(1)}(1/\delta)$ steps to be in an optimum solution with probability $1-\delta$, where n denotes the total number of tasks. The auxiliary computations are of polynomial complexity. Since the model cannot be approximated arbitrarily closely in the general case (unless P = NP), the approach can be used to obtain approximation algorithms that work well in the average case.

Keywords: Flow shop scheduling; Simulated annealing; Logarithmic cooling schedule; Convergence

1. Introduction

In the flow shop scheduling problem, l jobs have to be processed on m different machines. Each job consists of a sequence of tasks that have to be processed during an

☆ Research partially supported by the Strategic Research Program at the Chinese University of Hong Kong under Grant No. SRP 9505, by a Hong Kong Government RGC Earmarked Grant, Ref. No. CUHK 4367/99E, and by the HK-Germany Joint Research Scheme under Grant No. D/9800710.

* Corresponding author.

E-mail address: kathleen@first.gmd.de (K. Steinhöfel).

[1] On leave from IBM T.J. Watson Research Center, P.O. Box 218, Yorktown Heights, NY, USA.

0304-3975/02/$ - see front matter
PII: S0304-3975(01)00293-6

uninterrupted time period of a given length. The order in which each job is processed by the machines is the same for all jobs. A schedule is an allocation of the tasks to time intervals on the machines and the aim is to find a schedule that minimizes the overall completion time which is called the makespan.

Flow shop scheduling has long been identified as having a number of important practical applications. Baumgärtel addresses in [4] the flow shop problem in order to deal with the planning of material flow in car plants. His approach was applied to the logistics for the Mercedes Benz automobile. A more general overview on scheduling algorithms and their application is given in [1,8,12,13]. The NP-hardness of the general problem setting with $m \geqslant 3$ was shown by Garey, Johnson, and Sethi [9] in 1976. The existence of a polynomial approximation scheme for the flow shop scheduling problem with an arbitrary fixed number of machines is demonstrated by Hall in [11]. A recent work of Williamson et al., constitutes theoretical evidence that the general problem, which is considered in the present paper, is hard to solve even approximately. They proved that finding a schedule that is shorter than $\frac{5}{4}$ times the optimum is NP-hard [15].

We are concentrating on the convergence analysis of simulated annealing-based algorithms which employ a logarithmic cooling schedule. The algorithms utilize a simple neighborhood which is reversible and ensures a priori that transitions always result in a feasible solution. The neighborhood relation determines a landscape of the objective function over the configuration space $\mathscr{F}$ of feasible solutions of a given flow shop scheduling problem. Let $\mathbf{a}_S(k)$ denote the probability to obtain the schedule $S \in \mathscr{F}$ after k steps of a logarithmic cooling schedule $c(k) = \Gamma / \ln(k+2)$. The problem is to find a lower bound for k such that $\sum_{S \in \mathscr{F}_{\min}} \mathbf{a}_S(k) > 1 - \delta$ for schedules S minimizing the makespan. The general framework of logarithmic cooling schedules has been studied intensely, e.g., by Hajek [10] and Catoni [6, 7]. Hajek proved a lower bound on Γ that ensures the convergence to optimum solutions. Catoni investigated the speed of convergence and obtained upper bounds for the probability to be in a non-optimum state of the type $O(k^{-\alpha})$. Parameters of the optimization problem like the number of configurations are treated as constants and α depends on the structure of the associated energy landscape. In our approach, we tried to find upper bounds which are independent of structural properties of the configuration space.

Our convergence result, i.e., the lower bound on the number of steps k that ensure $\sum_{S \in \mathscr{F}_{\min}} \mathbf{a}_S(k) > 1 - \delta$, is based on a very detailed analysis of transition probabilities between neighboring elements of the configuration space $\mathscr{F}$. We obtain a run-time of $n^{O(\Gamma)} + \log^{O(1)}(1/\delta)$ to have with probability $1 - \delta$ a schedule with the minimum value of the makespan, where Γ is a parameter of the energy landscape characterizing the escape depth from local minima.

To our knowledge, logarithmic simulated annealing has not been applied previously to flow shop scheduling. Thus, we cannot present a comparison to results from the literature. In [15, 16], we used simulated annealing in computational experiments on job shop scheduling benchmark problems. For example, for one of the unsolved Yamada/Nakano instances we succeeded to improve the upper bound of the makespan. However, in job shop scheduling, the configuration space induced by the disjunctive

graph model and appropriate neighborhood relation is not reversible (see the remarks after Definition 1). Therefore, our convergence result cannot be extended in a straightforward way to job shop scheduling.

2. The flow shop problem

The flow shop scheduling problem can be formalized as follows. There are a set $\mathscr{J}$ of l jobs and a set $\mathscr{M}$ of m machines. Each job has exactly one task to be processed on each machine. Therefore, we have $n := lm$ tasks each with a given processing time $p(t) \in \mathbb{N}$. There is a binary relation R on the set of tasks $\mathscr{T}$ that decomposes $\mathscr{T}$ into chains corresponding to the jobs. The binary relation, which represents *precedences* between the tasks is defined as follows: For every $t \in \mathscr{T}$ there exists at most one t' such that $(t,t') \in R$. If $(t,t') \in R$, then $J(t) = J(t')$ and there is no $x \notin \{t,t'\}$ such that $(t,x) \in R$ or $(x,t') \in R$. For any $(v,w) \in R$, v has to be performed before w. R induces a total ordering of the tasks belonging to the same job. There exist no precedences between tasks of different jobs. Clearly, if $(v,w) \in R$ then $M(v) \neq M(w)$. The order in which a job passes all machines is the same for all jobs.

In Table 1 we give a small example of a flow shop problem. It consists of four jobs that need to be processed on four machines. As the *task number* of a task t we will denote the number of tasks preceding t within its job. We can therefore assume that all tasks with task number i are processed on machine M_i. A schedule is a function $S : \mathscr{T} \to \mathbb{N} \cup \{0\}$ that defines for each task t a starting time $S(t)$.

The *length*, respectively, the *makespan* of a schedule S is defined by

$$\lambda(S) := \max_{v \in \mathscr{T}} (S(v) + p(v)), \tag{1}$$

i.e., the earliest time at which all tasks are completed. The problem is to find an *optimum* schedule, that is feasible and of minimum length.

A flow shop scheduling problem can be represented by a *disjunctive graph*, a model introduced by Roy and Sussmann in [14]. As shown in Fig. 1 the disjunctive graph is a graph $G = (V, A, E, \mu)$, which is defined as follows:

$$\begin{aligned} V &= \mathscr{T} \cup \{I, O\}, \\ A &= \{[v,w] \mid v,w \in \mathscr{T}, (v,w) \in R\} \cup \{[I,w] \mid w \in \mathscr{T}, \nexists v \in \mathscr{T} : (v,w) \in R\} \cup \\ &\quad \{[v,O] \mid v \in \mathscr{T}, \nexists w \in \mathscr{T} : (v,w) \in R\}, \end{aligned}$$

Table 1
Example of a flow shop instance with 4 jobs and 4 machines

	T_0	T_1	T_2	T_3
J_0	$M_0/7$	$M_1/4$	$M_2/1$	$M_3/4$
J_1	$M_0/8$	$M_1/1$	$M_2/6$	$M_3/3$
J_2	$M_0/1$	$M_1/4$	$M_2/2$	$M_3/5$
J_3	$M_0/4$	$M_1/4$	$M_2/3$	$M_3/1$

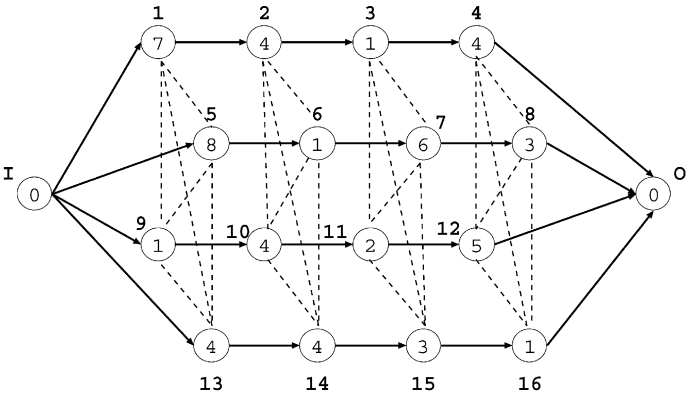

Fig. 1. The disjunctive graph G of the example from Table 1.

$$E = \{\{v,w\} \mid v,w \in \mathscr{T}, v \neq w, M(v) = M(w)\},$$
$$\mu : V \to \mathbb{N}.$$

The vertices in V represent the tasks. In addition, there are a source (I) and a sink (O) which are two dummy vertices. All vertices in V are weighted. The weight of a vertex $\mu(v)$ is given by the processing time $p(v)$, $\mu(v) := p(v)$, $(\mu(I) = \mu(O) = 0)$. The arcs in A represent the given precedences between the tasks. The edges in E represent the machine capacity constraints, i.e., $\{v,w\} \in E$ with $v,w \in \mathscr{T}$ and $M(v) = M(w)$ denotes the disjunctive constraint and the two ways to settle the disjunction correspond to the two possible orientations of $\{v,w\}$. The source I has arcs emanating to all first tasks of the jobs and the sink O has arcs coming from all final tasks of jobs.

An *orientation* on E is a function $\Omega : E \to \mathscr{T} \times \mathscr{T}$ such that $\Omega(\{v,w\}) \in \{\langle v,w\rangle, \langle w,v\rangle\}$ for each $\{v,w\} \in E$. A schedule is *feasible* if the corresponding orientation Ω on E $(\Omega(E) = \{\Omega(e) \mid e \in E\})$ results in a directed graph (called digraph) $D := G' = (V, A, E, \mu, \Omega(E))$ which is acyclic.

A *path* P from x_i to x_j, $i,j \in \mathbb{N}, i < j : x_i, x_j \in V$ of the digraph D is a sequence of vertices $(x_i, x_{i+1}, \ldots, x_j) \in V$ such that for all $i \leqslant k < j$, $[x_k, x_{k+1}] \in A$ or $\langle x_k, x_{k+1}\rangle \in \Omega(E)$.

The length of a path $P(x_i, x_j)$ is defined by the sum of the weights of all vertices in P: $\lambda(P(x_i, x_j)) = \sum_{k=i}^{j} \mu(x_k)$. The makespan of a feasible schedule is determined by the length of a longest path (i.e., a critical path) in the digraph D. The problem of minimizing the makespan therefore can be reduced to finding an orientation Ω on E that minimizes the length of $\lambda(P_{\max})$.

Table 2 shows a benchmark instance of flow shop scheduling. It was introduced by Carlier [5] and is called *car*7 since it consists of seven jobs on seven machines. In order to illustrate a feasible schedule and that an optimum makespan does not imply that the order of jobs on each machine is the same we give in Fig. 2 an example of an optimum solution of *car*7. The bar graph used in Fig. 2 is called Gantt chart named

Table 2
A flow shop instance with 7 jobs

	T_0	T_1	T_2	T_3	T_4	T_5	T_6
J_0	$M_0/692$	$M_1/310$	$M_2/832$	$M_3/630$	$M_4/258$	$M_5/147$	$M_6/255$
J_1	$M_0/581$	$M_1/582$	$M_2/14$	$M_3/214$	$M_4/147$	$M_5/753$	$M_6/806$
J_2	$M_0/475$	$M_1/475$	$M_2/785$	$M_3/578$	$M_4/852$	$M_5/2$	$M_6/699$
J_3	$M_0/23$	$M_1/196$	$M_2/696$	$M_3/214$	$M_4/586$	$M_5/356$	$M_6/877$
J_4	$M_0/158$	$M_1/325$	$M_2/530$	$M_3/785$	$M_4/325$	$M_5/565$	$M_6/412$
J_5	$M_0/796$	$M_1/874$	$M_2/214$	$M_3/963$	$M_4/896$	$M_5/898$	$M_6/302$
J_6	$M_0/542$	$M_1/205$	$M_2/578$	$M_3/578$	$M_4/325$	$M_5/800$	$M_6/120$

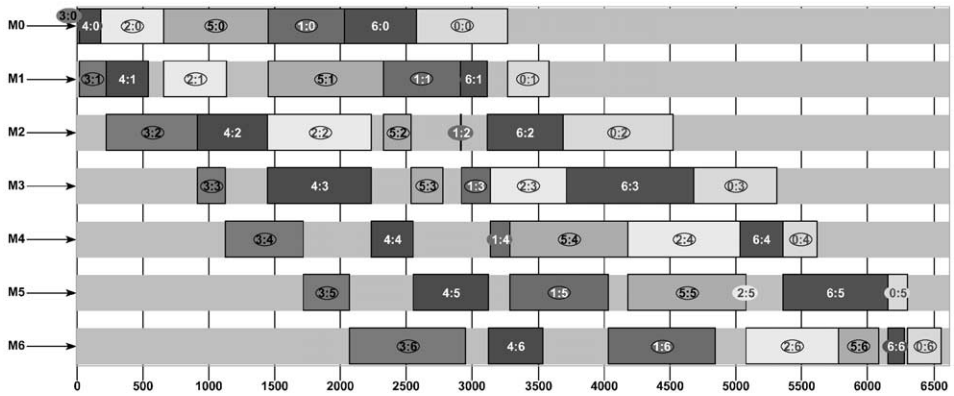

Fig. 2. A Gantt chart representing a feasible schedule of the example in Table 2.

after H. Gantt. It is a standard scheme to visualize schedules. Each color indicates a job and each row symbolizes a machine. The tasks are represented as boxes and labeled with the job number (jn) and the task number (tn) such that ($jn:tn$). The x-axis shows the time and therefore, the width of a box equals the processing time of the task. At time 6558 all tasks are completed, the makespan equals 6558 which is an optimum solution of *car*7 [5].

3. Basic definitions

Simulated annealing algorithms are acting within a configuration space in accordance with a certain neighborhood structure or a set of transition rules, where the particular steps are controlled by the value of an objective function. The configuration space, i.e., the set of feasible solutions of a given problem instance, is denoted by $\mathscr{F}$. For all instances, the number of tasks of each job equals the number of machines and each job has precisely one operation on each machine. Therefore, the size of $\mathscr{F}$ can be upper bounded in the following way. In the disjunctive graph G there are at most

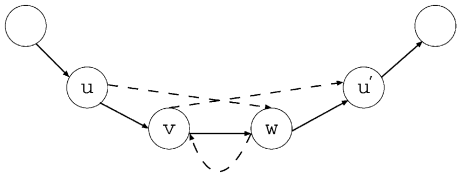

Fig. 3. The transition according to η.

$l!$ possible orientations to process l tasks on a single machine. Hence, we have $|\mathscr{F}| \leqslant (l!)^m$.

To describe the neighborhood of a solution $S \in \mathscr{F}$, we define a *neighborhood function* $\eta : \mathscr{F} \to \wp(\mathscr{F})$. The neighborhood of S is given by $\eta(S) \subseteq \mathscr{F}$, and each solution in $\eta(S)$ is called a neighbor of S. Van Laarhoven et al. [17] propose a neighborhood function η_L for solving job shop scheduling problems which is based on interchanging two adjacent tasks of a block. A *block* is a maximal sequence of adjacent tasks that are processed on the same machine and do belong to a longest path. We will use their neighborhood function with the extension that we allow changing the orientation of an arbitrary arc which connects two tasks on the same machine:

(i) Choosing two vertices v and w such that $M(v) = M(w) = k$ with $e = \langle v, w\rangle \in \Omega(E)$;
(ii) Reversing the order of e such that the resulting arc $e' \in \Omega'(E)$ is $\langle w, v\rangle$;
(iii) If there exists an arc $\langle u, v\rangle$ such that $v \neq u, M(u) = k$, then replace the arc $\langle u, v\rangle$ by $\langle u, w\rangle$;
(iv) If there exists an arc $\langle w, u'\rangle$ such that $w \neq u', M(u') = k$, then replace the arc $\langle w, u'\rangle$ by $\langle v, u'\rangle$.

Fig. 3 illustrates our transition rules. Thus, the neighborhood structure is characterized by

Definition 1. The schedule S' is a neighbor of S, $S' \in \eta(S)$, if S' can be obtained by the transition rules $1-4$ or $S' = S$.

Our choice is motivated by two facts:

- In contrast to the job shop scheduling the transition rules do guarantee for the flow shop a priori that the resulting schedule is feasible, i.e., that the corresponding digraph is acyclic.
- The extension of allowing to reverse the orientation of an arbitrary arc leads to an important property of the neighborhood function, namely reversibility.

Thus, the neighborhood structure is such that the algorithm visits only digraphs corresponding to feasible solutions and is equipped with a symmetry property which is required by our convergence analysis.

Lemma 1. *Suppose that $e = \langle v, w\rangle \in \Omega(E)$ is an arbitrary arc of an acyclic digraph D. Let D' be the digraph obtained from D by reversing the arc e. Then D' is also acyclic.*

Proof. Suppose D' is cyclic. Any arc $e=\langle v,w\rangle$ of D that is allowed to be reversed connects tasks which are processed on the same machine only. Because D is acyclic, the arc $\langle w,v\rangle$ is part of the cycle in D'. Consequently, there is a path $P=(v,x_1,x_2,\ldots,x_i,w)$ in D'. Since w is processed before v on machine M_k at least two arcs of the path P are connecting vertices of the same job. From the definition of the flow shop problem that implies at least two vertices have a task number greater than k. Neither within a job nor within a machine there is an arc $\langle y,z\rangle$ such that the task number of y is greater than the task number of z. This contradicts that the path P exists in D'. Hence, D' is acyclic. □

As already mentioned in Section 2, the objective is to minimize the makespan of feasible schedules. Hence, we define $\mathscr{L}(S):=\lambda(P_{\max})$, where $P_{\max}$ is a longest path in $D(S)$. Furthermore, we set

$$\mathscr{F}_{\min}:=\{S\mid S\in\mathscr{F}\text{ and }\forall S'(S'\in\mathscr{F}\to\mathscr{L}(S')\geqslant\mathscr{L}(S))\}. \tag{2}$$

For the special case of η_{L}, Van Laarhoven et al. have proved the following.

Theorem 1 (Van Laarhoven et al. [17]). *For each schedule $S\notin\mathscr{F}_{\min}$, there exists a finite sequence of transitions leading from S to an element of $\mathscr{F}_{\min}$.*

The probability of generating a solution S' from S can be expressed by

$$G[S,S']:=\begin{cases}1/|\eta| & \text{if } S'\in\eta(S),\\ 0 & \text{otherwise,}\end{cases} \tag{3}$$

with $|\eta|\leqslant n-m+1$ which follows from Definition 1.

The acceptance probability $A[S,S']$, $S'\in\eta(S)\subseteq\mathscr{F}$, is given by

$$A[S,S']:=\begin{cases}1 & \text{if } \mathscr{L}(S')-\mathscr{L}(S)\leqslant 0,\\ \mathrm{e}^{-(\mathscr{L}(S')-\mathscr{L}(S))/c} & \text{otherwise,}\end{cases} \tag{4}$$

where c is a control parameter having the interpretation of a *temperature* in annealing procedures. Finally, the probability of performing the transition between S and S', $S,S'\in\mathscr{F}$, is defined by

$$\mathbf{Pr}\{S\to S'\}=\begin{cases}G[S,S']A[S,S'] & \text{if } S'\neq S,\\ 1-\sum\limits_{Q\neq S}G[S,Q]A[S,Q] & \text{otherwise.}\end{cases} \tag{5}$$

Let $\mathbf{a}_S(k)$ denote the probability of being in the configuration S after k steps performed for the same value of c. The probability $\mathbf{a}_S(k)$ can be calculated in accordance

with

$$\mathbf{a}_S(k) := \sum_Q \mathbf{a}_Q(k-1)\ \mathbf{Pr}\{Q \to S\}. \tag{6}$$

The recursive application of (6) defines a Markov chain of probabilities $\mathbf{a}_S(k)$. If the parameter $c = c(k)$ is a constant c, the chain is said to be a *homogeneous* Markov chain; otherwise, if $c(k)$ is lowered at any step, the sequence of probability vectors $\mathbf{a}(k)$ is an *inhomogeneous* Markov chain.

We consider a cooling schedule which defines a special type of inhomogeneous Markov chains. For this cooling schedule, the value $c(k)$ changes in accordance with

$$c(k) = \frac{\Gamma}{\ln(k+2)}, \quad k = 0, 1, \ldots\ . \tag{7}$$

The choice of $c(k)$ is motivated by Hajek's Theorem [10] on logarithmic cooling schedules for inhomogeneous Markov chains. If there exists $S_0, S_1, \ldots, S_r \in \mathscr{F}$ ($S_0 = S \wedge S_r = S'$) such that $G[S_u, S_{u+1}] > 0, u = 0, 1, \ldots, (r-1)$ and $\mathscr{Z}(S_u) \leqslant h$, for all $u = 0, 1, \ldots, r$, we denote $height(S \Rightarrow S') \leqslant h$. The schedule S is a *local minimum*, if $S \in \mathscr{F} \setminus \mathscr{F}_{\min}$ and $\mathscr{Z}(S') > \mathscr{Z}(S)$ for all $S' \in \eta_L(S) \setminus S$. By $depth(S_{\min})$ we denote the smallest h such that there exists a $S' \in \mathscr{F}$, where $\mathscr{Z}(S') < \mathscr{Z}(S_{\min})$, which is reachable at height $\mathscr{Z}(S_{\min}) + h$.

The following convergence property has been proved by Hajek.

Theorem 2 (Hajek [10]). *Given a configuration space $\mathscr{C}$ and a cooling schedule defined by*

$$c(k) = \frac{\Gamma}{\ln(k+2)}, \quad k = 0, 1, \ldots,$$

the asymptotic convergence $\sum_{H \in \mathscr{C}} \mathbf{a}_H(k) \to_{k \to \infty} 1$ of the stochastic algorithm, which is based on (2), (4), *and* (5), *is guaranteed if and only if*

(i) $\forall H, H' \in \mathscr{C}\ \exists H_0, H_1, \ldots, H_r \in \mathscr{C} (H_0 = H \wedge H_r = H')$: $G[H_u, H_{u+1}] > 0$, $l = 0, 1, \ldots, (r-1)$;
(ii) $\forall h : height(H \Rightarrow H') \leqslant h \Leftrightarrow height(H' \Rightarrow H) \leqslant h$;
(iii) $\Gamma \geqslant \max_{H_{\min}} depth(H_{\min})$.

Condition (i) expresses the connectivity of the configuration space. As already mentioned above, with the choice of our neighborhood relation we can guarantee the mutual reachability of schedules. Therefore, Hajek's Theorem can be applied to our configuration space $\mathscr{F}$ with the neighborhood relation η.

Before we perform the convergence analysis of the logarithmic cooling schedule defined in (7), we point out some properties of the configuration space and the neighborhood function. Let S and S' be feasible schedules and $S' \in \eta(S)$. To obtain S' from S, we chose the arc $e = \langle v, w \rangle$ with $M(v) = M(w)$.

If $\mathscr{Z}(S) < \mathscr{Z}(S')$, then only a path containing one of the selected vertices v, w can determine the new makespan after the transition move. It can be shown that all paths

whose length increases contain the edge $e' = \langle w, v \rangle$. Therefore, we have the following upper bound.

Lemma 2. *The increase of the objective function $\Delta\mathscr{Z}$ in a single step according to η $(S \to_\eta S')$ can be upper by $p(v) + p(w)$.*

The reversibility of the neighborhood function implies for the maximum distance of neighbors $S' \in \eta(S)$ to $\mathscr{F}_{\min}$ in relation to S itself: If the minimum number of transitions to reach from S an optimum element is N, then for any $S' \in \eta(S)$ the minimum number of transitions is at most $N + 1$.

4. Convergence analysis

Our convergence results will be derived from a careful analysis of the "exchange of probabilities" among feasible solutions which belong to adjacent distance levels to optimum schedules, i.e., in addition to the value of the objective function, the elements of the configuration space are further distinguished by the minimal number of transitions required to reach an optimum schedule. We first introduce a recurrent formula for the expansion of probabilities and then we prove the main result on the convergence rate which relates properties of the configuration space to the speed of convergence. Throughout the section we employ the fact that for a proper choice of Γ the logarithmic cooling schedule leads to an optimum solution. Further, we employ a method that has been developed in the context of equilibrium computations [2, 3].

To express the relation between S and S' according to their value of the objective function we will use $<_{\mathscr{Z}}$, $>_{\mathscr{Z}}$, and $=_{\mathscr{Z}}$ to simplify the expressions

$$\begin{aligned} &S <_{\mathscr{Z}} S' \text{ instead of } \quad S' \in \eta(S) \wedge (\mathscr{Z}(S) < \mathscr{Z}(S')),\\ &S >_{\mathscr{Z}} S' \text{ instead of } \quad S' \in \eta(S) \wedge (\mathscr{Z}(S) > \mathscr{Z}(S')),\\ &S =_{\mathscr{Z}} S' \text{ instead of } \quad S \neq S' \wedge S' \in \eta(S) \wedge (\mathscr{Z}(S) = \mathscr{Z}(S')). \end{aligned}$$

Furthermore, we denote

$$p(S) := |\{S <_{\mathscr{Z}} S'\}|, \quad q(S) := |\{S =_{\mathscr{Z}} S'\}|, \quad r(S) := |\{S >_{\mathscr{Z}} S'\}|.$$

These notations imply

$$p(S) + q(S) + r(S) = |\eta(S)| - 1 = ml - m - 1. \tag{8}$$

The equation is valid because there are m $(l-1)$ arcs which are allowed to be switched and S belongs to its own neighborhood. Therefore, the size of the neighborhood is independent of the particular schedule S, and we set $n' := ml - m$.

Now, we analyze the probability $\mathbf{a}_S(k)$ to be in the schedule $S \in \mathscr{F}$ after k steps of the logarithmic cooling schedule defined in Eq. (7), and we use the notation

$$\frac{1}{(k+2)^{(\mathscr{Z}(S)-\mathscr{Z}(S'))/\Gamma}} = \mathrm{e}^{-(\mathscr{Z}(S)-\mathscr{Z}(S'))/c(k)}, \quad k \geqslant 0. \tag{9}$$

By using (3)–(5), one obtains from (6) by straightforward calculations

$$\mathbf{a}_S(k) = \mathbf{a}_S(k-1)\left(\frac{p(S)+1}{n'} - \sum_{\substack{i=1 \\ S <_{\mathscr{Z}} S_i}}^{p(S)} \frac{1}{n'}\frac{1}{(k+1)^{(\mathscr{Z}(S_i)-\mathscr{Z}(S))/\Gamma}}\right)$$

$$+ \sum_{\substack{i=1 \\ S_i \geqslant_{\mathscr{Z}} S}}^{p(S)+q(S)} \frac{\mathbf{a}_{S_i}(k-1)}{n'} + \sum_{\substack{j=1 \\ S_j <_{\mathscr{Z}} S}}^{r(S)} \frac{\mathbf{a}_{S_j}(k-1)}{n'}\frac{1}{(k+1)^{(\mathscr{Z}(S)-\mathscr{Z}(S_j))/\Gamma}}. \tag{10}$$

The representation (expansion) will be used in the following as the main relation reducing $\mathbf{a}_S(k)$ to probabilities from previous steps.

We introduce the following partition of the set of schedules with respect to the value of the objective function:

$$L_0 := \mathscr{F}_{\min}$$

and

$$L_{h+1} := \left\{ S : S \in \mathscr{F} \wedge \forall S' \left(S' \in \mathscr{F} \setminus \bigcup_{i=0}^{h} L_i \rightarrow \mathscr{Z}(S') \geqslant \mathscr{Z}(S) \right) \right\}.$$

The highest level within $\mathscr{F}$ is denoted by $L_{h_{\max}}$. Given $S \in \mathscr{F}$, we further denote by $\mathscr{W}_{\min}(S) := [S, S_{k-1}, \ldots, S']$ a shortest sequence of transitions from S to $\mathscr{F}_{\min}$, i.e., $S' \in \mathscr{F}_{\min}$. Thus, we have for the distance $d(S) := \mathit{length}(\mathscr{W}_{\min}(S))$. We introduce another partition of $\mathscr{F}$ with respect to $d(S)$:

$$S \in D_i \Leftrightarrow d(S) = i \geqslant 0, \quad \text{and} \quad \mathscr{D}_{\mathbf{s}} = \bigcup_{i=1}^{\mathbf{s}-1} D_i, \ \text{i.e.,} \ \mathscr{F} = \mathscr{D}_{\mathbf{s}}.$$

Thus, we distinguish between distance levels D_i related to the minimal number of transitions required to reach an optimal schedule from $\mathscr{F}_{\min}$ and the levels L_h which are defined by the objective function. By definition, we have $D_0 := L_0 = \mathscr{F}_{\min}$. We will use the following abbreviations:

$$f(S', S, t) := \frac{1}{(k+2-t)^{(\mathscr{Z}(S')-\mathscr{Z}(S))/\Gamma}} \tag{11}$$

and

$$K_S(k-t) := \frac{p(S)+1}{n'} - \sum_{\substack{i=1 \\ S <_{\mathscr{Z}} S_i}}^{p(S)} \frac{1}{n'}(k+2-t)^{-(\mathscr{Z}(S_i)-\mathscr{Z}(S))/\Gamma}. \tag{12}$$

We are going backwards from the kth step and expanding $\mathbf{a}_S(k)$ in accordance with (10). Our aim is to find a close upper bound for the value $\sum_{S \notin D_0} \mathbf{a}_S(k)$ in terms of probabilities from previous steps.

During the expansion (10) of $\mathbf{a}_S(k)$, $S \notin D_0$, terms according to S are generated as well as according to all neighbors S' of S. Some terms generated by the expansion of

S contain the factor $\mathbf{a}_{S'}(k-1)$ and can therefore be summarized with terms generated by the expansion of S'. However, it is important to distinguish between elements from D_1 and elements from D_i, $i \geqslant 2$. For all $S \notin D_1$, we obtain the following:

$$\mathbf{a}_S(k-1)\left(\frac{p(S)+1+q(S)+r(S)}{n'} - \sum_{\substack{i=1 \\ S <_{\mathscr{Z}} S_i}}^{p(S)} \frac{1}{n'} \frac{1}{(k+1)^{(\mathscr{Z}(S_i)-\mathscr{Z}(S))/\Gamma}} + \sum_{\substack{i=1 \\ S <_{\mathscr{Z}} S_i}}^{p(S)} \frac{1}{n'} \frac{1}{(k+1)^{(\mathscr{Z}(S_i)-\mathscr{Z}(S))/\Gamma}}\right) = \mathbf{a}_S(k-1). \tag{13}$$

In case of $S \in D_1$, some neighbors S' of S are elements of D_0 and do not generate the terms related to $S >_{\mathscr{Z}} S'$ because the $\mathbf{a}_{S'}(k)$ are not expanded since they are not present in the sum $\sum_{S \notin D_0} \mathbf{a}_S(k)$. Therefore, $r'(S) \leqslant r(S)$ many terms are missing for $S \in D_1$ and the following arithmetic term is generated:

$$\mathbf{a}_S(k-1)\left(1 - \frac{r'(S)}{n'}\right), \tag{14}$$

where $r'(S) := |\{S' : S' \in \eta(S) \wedge S' \in D_0\}|$. On the other hand, the expansion of $\mathbf{a}_S(k)$ generates terms related to $S' \in D_0$ with $S' <_{\mathscr{Z}} S$ and containing $\mathbf{a}_{S'}(k-1)$ as a factor. Those terms are not canceled by expansions of $\mathbf{a}_{S'}(k)$. All $S \in D_1$ therefore generate the following term:

$$\sum_{\substack{j=1 \\ S_j \in D_0 \cap \eta(S)}}^{r'(S)} \frac{\mathbf{a}_{S_j}(k-1)}{n'} \frac{1}{(k+1)^{(\mathscr{Z}(S)-\mathscr{Z}(S_j))/\Gamma}}. \tag{15}$$

Now, we consider the entire sum and take the negative product $\mathbf{a}_S(k) r'(S)/n'$ separately. By using the abbreviations introduced in Eq. (12) we derive the following lemma.

Lemma 3. *After one step of the expansion of* $\sum_{S \notin D_0} \mathbf{a}_S(k)$, *the sum can be represented by*

$$\sum_{S \notin D_0} \mathbf{a}_S(k) = \sum_{S \notin D_0} \mathbf{a}_S(k-1) - \sum_{S \in D_1} \frac{r'(S)}{n'} \mathbf{a}_S(k-1) + \sum_{S \in D_1} \sum_{\substack{j=1 \\ S_j \in D_0 \cap \eta(S)}}^{r'(S)} \frac{f(S,S_j,1)}{n'} \mathbf{a}_{S_j}(k-1).$$

The diminishing factor $(1 - r'(S)/n')$ appears by definition for all elements of D_1. At subsequent reduction steps, the factor is "transmitted" successively to all probabilities from higher distance levels D_i because any element of D_i has at least one neighbor from D_{i-1}. The main task is now to analyze how this diminishing factor changes, if

it is propagated to the next higher distance level. We denote

$$\sum_{S\notin D_0} \mathbf{a}_S(k) = \sum_{S\notin D_0} \mu(S,t)\mathbf{a}_S(k-t) + \sum_{S'\in D_0} \mu(S',t)\mathbf{a}_{S'}(k-t), \tag{16}$$

i.e., the coefficients $\mu(S,t)$ and $\mu(S',t)$ are the factors at probabilities after t steps of an expansion of $\sum_{S\notin D_0} \mathbf{a}_S(k)$. Hence, for $S\in D_1$ we have $\mu(S,1)=1-r'(S)/n'$, and $\mu(S,1)=1$ for the remaining $S\in\mathscr{D}_\mathbf{s}\backslash(D_0\cup D_1)$. For $S'\in D_0$ we have from Lemma 3

$$\mu(S',1) = \sum_{\substack{i=1 \\ S_i\in D_1\wedge S'\in\eta(S_i)}}^{p(S')} \frac{f(S_i,S',1)}{n'}. \tag{17}$$

Starting from step $(k-1)$, the generated probabilities $\mathbf{a}_{S'}(k-u)$ are expanded in the same way as all other probabilities. We set $\mu(S,j):=1-v(S,j)$ because we are mainly interested in the convergence $\mu(S,j)\to 0$. We perform an inductive step from $(k-t+1)$ to $(k-t)$ and obtain for $t\geqslant 2$:

Lemma 4. *The following recurrent relation is valid for the coefficients* $v(S,t)$, $t\geqslant 2$:

$$v(S,t) = v(S,t-1)K_S(k-t) + \sum_{S\geqslant_{\mathscr{T}} S'} \frac{v(S',t-1)}{n'} + \sum_{S<_{\mathscr{T}} S''} \frac{v(S'',t-1)}{n'} f(S'',S,t).$$

Furthermore, for the three special cases $S\in D_j$, $j>t$, $S\in D_1, t=1$, *and* $S\in D_0$, $t=1$ *we have,* $v(S,t)=0$, $v(S,t)=r'(S)/n'$, *and* $v(S,t)=1-\sum_{j=1}^{p(S)} f(S_j,S,1)/n'$, *with* $S_j\in D_1 \wedge S\in\eta(S_j)$, *respectively.*

Exactly the same structure of the equation is valid for $\mu(S,t)$ which will be used for elements of D_0 only because these elements are not present in the original sum $\sum_{S\notin D_0} \mathbf{a}_S(k)$. Now, any $v(S,t)$ and $\mu(S,t)$ is expressed by a sum $\sum_u T_u$ of arithmetic terms. We consider in more details the terms associated with elements $S^{\mathbf{0}}$ of D_0 and $S^{\mathbf{1}}$ of D_1. We assume a representation $\mu(S^{\mathbf{0}},t-1)=\sum T(S^{\mathbf{0}})$, and $v(S,t-1)=\sum T(S)$, $S\notin D_0$.

If we consider $r'(S^{\mathbf{1}})/n'$ and $\sum_{S^{\mathbf{0}}<_{\mathscr{T}} S^{\mathbf{1}}} f(S^{\mathbf{1}},S^{\mathbf{0}},t)/n'$ separately, the difficulties arising from the definition $v(S,t):=1-\mu(S,t)$ can be avoided, i.e., we have to take into account only changing signs of terms during the transmission from D_1 to D_0 and vice versa.

Definition 2. The two expressions $r'(S^{\mathbf{1}})/n'$, and $\sum_{S^{\mathbf{0}}<_{\mathscr{T}} S^{\mathbf{1}}} f(S^{\mathbf{1}},S^{\mathbf{0}},t)/n'$, are called *source terms* of $v(S^{\mathbf{1}},t)$ and $\mu(S^{\mathbf{0}},t)$, respectively.

During an expansion of $\sum_{S\notin D_0} \mathbf{a}_S(k)$ backwards according to (13), the source terms are distributed permanently to higher distance levels D_j. Therefore, at higher distance levels the notion of a source term can be defined by an inductive step:

Definition 3. For all $S\in D_i,\ i>1$, any term which is generated according to the equation of Lemma 4 from a source term of $v(S',t-1)$, where $S'\in D_{i-1}\cap\eta(S)$, is said to be a source term of $v(S,t)$.

We introduce a counter $\mathbf{e}(T)$ to terms T which indicates the step at which the term has been generated from source terms. The value $\mathbf{e}(T)$ is called the rate of a term and we set $\mathbf{e}(T)=1$ for source terms T.

The value $\mathbf{e}(T)>1$ is assigned to terms related to D_0 and D_1 in a slightly different way compared to higher distance levels because at the first step the $S^{\mathbf{0}}$ do not participate in the expansion of $\sum_{S\notin D_0}\mathbf{a}_S(k)$. Furthermore, in the case of D_0 and D_1 we have to take into account the changing signs of terms which result from the simultaneous consideration of $v(S^{\mathbf{1}},t)$ (for D_1) and $\mu(S^{\mathbf{0}},t)$ (for D_0).

Definition 4. A term $T^{\mathbf{0}}$ is called a jth *rate term* of $\mu(S^{\mathbf{0}},t)$, $S^{\mathbf{0}}\in D_0$ and $j\geqslant 2$, if either $S^{\mathbf{0}}=-T$ and $\mathbf{e}(T)=j-1$ for some $v(S,t-1)$, $S\in D_1\cap\eta(S^{\mathbf{0}})$, or $\mathbf{e}(T^{\mathbf{0}})=j-1$ for some $\mu(S',t-1)$, $S'\in D_0\cap\eta(S^{\mathbf{0}})$.

A term T is called a jth *rate term* of $v(S,t)$, $S^{\mathbf{1}}\in D_1$ and $j\geqslant 2$, if $\mathbf{e}(T)=j-2$ for some $v(S',t-1)$, $S'\in D_2\cap\eta(S^{\mathbf{1}})$, $\mathbf{e}(T)=j-1$ for some $v(S',t-1)$, $S'\in D_1\cap\eta(S^{\mathbf{1}})$, or $T=-T'$ and $\mathbf{e}(T')=j-1$ for some $S^{\mathbf{0}}\in D_0\cap\eta(S^{\mathbf{1}})$ with respect to $\mu(S^{\mathbf{0}},t-1)$.

A term T is called a jth *rate term* of $v(S,t)$, $S\in D_i$ and $i,\ j\geqslant 2$, if $\mathbf{e}(T)=j-1$ for some $v(S',t-1)$, $S'\in D_{i+1}\cap\eta(S)$, $\mathbf{e}(T)=j-1$ for some $v(S',t-1)$, $S'\in D_i\cap\eta(S)$, or T is a jth rate term of $v(S,t-1)$ for some $S\in D_{i-1}$.

The classification of terms will be used for a partition of the summation over all terms which constitute particular values $v(S^{\mathbf{1}},t)$ and $\mu(S^{\mathbf{0}},t)$. Let $\mathscr{T}_j(S,t)$ be the set of jth rate arithmetic terms of $v(S^{\mathbf{1}},t)$ ($\mu(S^{\mathbf{0}},t)$) related to $S\in\mathscr{D}_{\mathbf{s}}$. We set

$$\mathbf{A}_j(S,t):=\sum_{T\in\mathscr{T}_j(S,t)}T. \tag{18}$$

The same notation is used in case of $S=S^{\mathbf{0}}\in D_0$ with respect to $\mu(S^{\mathbf{1}},t)$, and we obtain by induction

$$v(S,t)=\sum_{j=1}^{t-i+1}\mathbf{A}_j(S,t)\quad\text{and}\quad\mu(S^{\mathbf{0}},t)=\sum_{j=1}^{t}\mathbf{A}_j(S^{\mathbf{0}},t). \tag{19}$$

For $S\in D_i\neq D_1$, D_0 and $j\geqslant 2$ we obtain immediately from Lemma 4 and Definition 4

$$\mathbf{A}_j(S,t)=\mathbf{A}_{j-1}(S,t-1)K_S(k-t)+\sum_{\substack{S'\leqslant_{\mathscr{T}}S\\ S'\in D_{i+1}}}\frac{\mathbf{A}_{j-2}(S',t-1)}{n'}+\sum_{\substack{S''>_{\mathscr{T}}S\\ S''\in D_{i+1}}}\frac{\mathbf{A}_{j-2}(S'',t-1)}{n'}f(S'',S,t)$$

$$+\sum_{\substack{S'\leqslant_{\mathscr{T}} S\\ S'\in D_i}}\frac{\mathbf{A}_{j-1}(S',t-1)}{n'}+\sum_{\substack{S''>_{\mathscr{T}} S\\ S''\in D_i}}\frac{\mathbf{A}_{j-1}(S'',t-1)}{n'}f(S'',S,t)$$

$$+\sum_{\substack{S'\leqslant_{\mathscr{T}} S\\ S'\in D_{i-1}}}\frac{\mathbf{A}_{j}(S',t-1)}{n'}+\sum_{\substack{S''>_{\mathscr{T}} S\\ S''\in D_{i-1}}}\frac{\mathbf{A}_{j}(S'',t-1)}{n'}f(S'',S,t). \tag{20}$$

In case of $S\in D_1$ and $j\geqslant 2$, we have in accordance with Definition 4:

$$\begin{aligned}\mathbf{A}_j(S,t)&=\mathbf{A}_{j-1}(S,t-1)K_S(k-t)\\&\quad+\sum_{\substack{S'\leqslant_{\mathscr{T}} S\\ S'\in D_1}}\frac{\mathbf{A}_{j-1}(S',t-1)}{n'}+\sum_{\substack{S''>_{\mathscr{T}} S\\ S''\in D_1}}\frac{\mathbf{A}_{j-1}(S'',t-1)}{n'}f(S'',S,t)\\&\quad+\sum_{\substack{S''>_{\mathscr{T}} S\\ S''\in D_2}}\frac{\mathbf{A}_{j-2}(S'',t-1)}{n'}f(S'',S,t)\\&\quad-\sum_{S'\in D_0\cap\eta(S)}\frac{\mathbf{A}_{j-1}(S',t-1)}{n'}.\end{aligned} \tag{21}$$

Finally, the corresponding relation for $S^{\mathbf{0}}$ is given by

$$\mathbf{A}_j(S^{\mathbf{0}},t)=\mathbf{A}_{j-1}(S^{\mathbf{0}},t-1)K_{S^{\mathbf{0}}}(k-t)-\sum_{S>_{\mathscr{T}}S^{\mathbf{0}}}\frac{\mathbf{A}_{j-1}(S,t-1)}{n'}f(S,S^{\mathbf{0}},t). \tag{22}$$

We compare the computation of $v(S,t)$ (and $\mu(S',t)$) for two different values $t=k_1$ and k_2, i.e., $v(S,t)$ is calculated backwards from k_1 and k_2, respectively. To distinguish between $v(S,t)$ and related values, which are defined for different k_1 and k_2, we will use an additional upper index. At this point, we use again the representation (19) of $v(S,t)$ (and the corresponding equation for $\mu(S',t)$).

Lemma 5. *Given $k_2\geqslant k_1$ and $S\in D_i$, then*

$$\mathbf{A}_2^1(S,t)=\mathbf{A}_2^2(S,k_2-k_1+t)\quad if\ t\geqslant i+2.$$

We recall that our main goal is to upper bound the sum $\sum_{S\notin D_0}\mathbf{a}_S(k)$. When $\vec{\mathbf{a}}(0)$ denotes the initial probability distribution, we have from Eq. (16)

$$\begin{aligned}\left|\sum_{S\notin D_0}\mathbf{a}_S(k_1)-\sum_{S\notin D_0}\mathbf{a}_S(k_2)\right|&\leqslant\left|\sum_{S\notin D_0}(v(S,k_2)-v(S,k_1))\mathbf{a}_S(0)\right|\\&\quad+\left|\left(\sum_{S'\in D_0}\mu(S',k_2)-\sum_{S'\in D_0}\mu(S',k_1)\right)\mathbf{a}_{S'}(0)\right|.\end{aligned} \tag{23}$$

Any element of the initial distribution $\vec{\mathbf{a}}(0)$ is simply replaced by 1 and we obtain for the first part of the sum in accordance with Lemma 5

$$\sum_{S\notin D_0} |(v(S,k_2)-v(S,k_1))\mathbf{a}_S(0)| \leqslant \sum_{S\notin D_0} \sum_{j=k_1+1}^{k_2} |\mathbf{A}_j^2(S,k_2)|. \tag{24}$$

Now, we are going to estimate the values $|\mathbf{A}_j^2(S,k_2)|$. The relations in (21)–(22) are expanded until D_1 and/or D_0 have been reached for $t=1$. The computation of $\mathbf{A}_j(S,t)$ can be represented by a tree, where the leaves are terms $\mathbf{A}_1(S',1)$ for $S'\in D_1\cup D_0$. Starting from these values, we are calculating $\mathbf{A}_j(S,t)$ backwards. We distinguish between the following types of terms that are calculated at intermediate steps: first, we consider negative and positive terms separately. We concentrate on positive terms only because there is no significant difference between these two types (except for an additional factor $f(S_i,S',1)/n'$ arising from elements of D_0; see Lemma 4). Secondly, the terms from D_1 distributed upwards are separated from the terms passing through D_0 since the latter terms are multiplied by $f(S_i,S',t)$. When $k_1\geqslant n^{\mathrm{O}(\Gamma)}$, each of these multiplications generates for a significant fraction of $t\leqslant k_1$ a small factor $1/n^{\mathrm{O}(1)}$. Therefore, we consider in more details only the case when these multiplications are not guaranteed, i.e., when the terms are distributed to higher levels of the objective function.

We use two ways to represent upper bounds of the values $\mathbf{A}_j(S',v)$, $v\leqslant t$, for $S'\in D_1$ and $S'\notin D_1$, respectively. For $S'\in D_1$, we use the notation by a single value $P_v(v)$, whereas the terms at higher levels are represented by a sum of terms arising from terms $P_v(v)$ of D_1

$$P'_1(v)+P'_2(v)+\cdots+P'_{v-1}(v).$$

The $P'_u(v)$ indicate that the original $P_u(v)$ has been multiplied by factors related to the particular transition step, cf. 21) till (22). The factors from the same level of the neighborhood are taken together, e.g., we consider $d':=(p_{i-1}/(n'\cdot(k+2-u)^{\omega})+r_{i-1}/n')$ for the neighbors from the lower level D_{i-1}.

From (21)–(22) we obtain the recursive relation

$$P_v(v) \leqslant K'P_{v-1}(v)+d''d'P_{v-1}(v)+d''\sum_{l=1}^{v-2} P'_1(v). \tag{25}$$

The factor d' summarizes the transition to the higher level at the previous step and d'' represents an upper bound for $(p_{i+1}/(n'(k+2-u)^{\omega})+r_{i+1}/n')$. It is important to note that $K'<K_{S'}$ since $(p_0/(n'(k+2-u)^{\omega})+r_0/n')$ is missing because the terms passing through D_0 are considered separately.

The recursive application of (25) generates products consisting of the same number of factors. These factors are of two types: factors K' from self-transitions (but smaller than $K_{S'}$) and factors of the type $(p/(n'(k+2-u)^{\omega})+r/n')$. As in [15], the products can be enumerated by the number of self-transitions, steps to higher levels and transitions to the same or lower levels. For a fixed assignment of numbers to these types of transitions, we derive an upper bound for the products in the same way as in [15]. The resulting upper bound leads to

Lemma 6. *Given* $n^{O(n)} \geqslant k_2 \geqslant k_1 > n^{O(\Gamma)}$, *then*

$$\left| \sum_{S \notin D_0} (v(S,k_2) - v(S,k_1))\mathbf{a}_S(0) \right| < 2^{-k_1^{1/\beta}}$$

for a suitable constant $\beta > 1$.

The same relation can be derived for the $\mu(S', k_{1/2})$.

For $l = m$, the size of $\mathscr{F}$ is upper bounded by $n^{O(n)}$ (see Section 3). Thus, a value of $k_2 < n^{O(n)}$ allows a complete search of $\mathscr{F}$. We assume that such an upper bound is sufficient for a convergence according to Theorem 2. Now, we can immediately prove

Theorem 3. *When the convergence is guaranteed for* $k_2 \leqslant n^{O(n)}$, *then the condition*

$$k > n^{O(\Gamma)} + \log^{O(1)} \frac{1}{\delta}$$

implies for arbitrary initial probability distributions $\vec{\mathbf{a}}(0)$ *and* $\delta > 0$*:*

$$\sum_{S \notin D_0} \mathbf{a}_S(k) < \delta \quad \textit{and therefore,} \quad \sum_{S' \in D_0} \mathbf{a}_{S'}(k) > 1 - \delta.$$

Proof. We choose k in accordance with Lemma 6 and we consider

$$\begin{aligned} \sum_{S \notin D_0} \mathbf{a}_S(k) &= \sum_{S \notin D_0} (\mathbf{a}_S(k) - \mathbf{a}_S(k_2)) + \sum_{S \notin D_0} \mathbf{a}_S(k_2) \\ &= \sum_{S \notin D_0} (v(S,k_2) - v(S,k))\mathbf{a}_S(0) \\ &\quad + \sum_{S' \in D_0} (\mu(S',k) - \mu(S',k_2))\mathbf{a}_{S'}(0) + \sum_{S \notin D_0} \mathbf{a}_S(k_2). \end{aligned}$$

The value k_2 from Lemma 6 is larger but independent of $k_1 = k$, i.e., we can take a $k_2 > k$ such that $\sum_{S \notin D_0} \mathbf{a}_S(k_2) < \delta/3$. Here, we employ Theorems 1 and 2, i.e., if the constant Γ from (7) is sufficiently large, the inhomogeneous simulated annealing procedure defined by (3)–(5) tends to the global minimum of $\mathscr{Y}$ on $\mathscr{F}$. We obtain the stated inequality, if additionally both differences $\sum_{S \notin D_0} (v(S,k_2) - v(S,k))$ and $\sum_{S' \in D_0} (\mu(S',k) - \mu(S',k_2))$ are smaller than $\delta/3$. Lemma 6 implies that the condition on the differences is satisfied in case of $k_1^{1/\beta} > \log(3/\delta)$. □

For scheduling problems, it seems to be difficult to estimate a priori the values of Γ (in other applications of logarithmic simulated annealing, relatively tight bound of Γ can be provided, cf. [3]). However, the value of Γ can be estimated by computational experiments when the algorithm is applied to flow shop scheduling problems. We followed a similar approach when logarithmic simulated annealing was used for solving job shop scheduling benchmark problems [15, 16].

Acknowledgements

The authors would like to thank the referees for their careful reading of earlier versions of the paper and helpful suggestions for an improved presentation.

References

[1] E.H.L. Aarts, Local Search in Combinatorial Optimization, Wiley, New York, 1997.
[2] A. Albrecht, S.K. Cheung, K.S. Leung, C.K. Wong, Stochastic simulations of two-dimensional composite packings, J. Comput. Phys. 136 (2) (1997) 559–579.
[3] A. Albrecht, S.K. Cheung, K.S. Leung, C.K. Wong, On the convergence of inhomogeneous markov chains approximating equilibrium placements of flexible objects, Comput. Optim. Appl., to appear.
[4] H. Baumgärtel, Distributed constraint processing for production logistics, Practical Application of Constraint Technology, UK, 1997.
[5] J. Carlier, Ordonnancements a contraintes disjonctives, R.A.I.R.O, Oper. Res. 12 (1978) 333–351.
[6] O. Catoni, Rough large deviation estimates for simulated annealing: applications to exponential schedules, Ann. Probab. 20 (3) (1992) 1109–1146.
[7] O. Catoni, Metropolis, simulated annealing, and iterated energy transformation algorithms: theory and experiments, J. Complexity 12 (4) (1996) 595–623.
[8] P. Chretienne, E.G. Coffman Jr., J.K. Lenstra, Z. Liu, Scheduling Theory and its Applications, Wiley, New York, 1995.
[9] M.R. Garey, D.S. Johnson, R. Sethi, The complexity of flow shop and job shop scheduling, Math. Oper. Res. 1 (1976) 117–129.
[10] B. Hajek, Cooling schedules for optimal annealing, Math. Oper. Res. 13 (1988) 311–329.
[11] L.A. Hall, Approximability of flow shop scheduling, 36th Annu. Symp. on Foundations of Computer Science, Milwaukee, Wisconsin, 1995, pp. 82–91.
[12] E. Nowicki, C. Smutnicki, The flow shop with parallel machines: A tabu search approach, European J. Oper. Res. 106 (1998) 226–253.
[13] M. Pinedo, Scheduling: Theory, Algorithms, and Systems, Prentice-Hall International Series in Industrial and Systems Engineering, Prentice-Hall, Englewood Cliffs, NJ, 1995.
[14] B. Roy, B. Sussmann, Les problèmes d'Ordonnancement avec Constraints Disjonctives, Note DS No. 9 bis. SEMA, 1964.
[15] K. Steinhöfel, A. Albrecht, C.K. Wong, On various cooling schedules for simulated annealing applied to the job shop problem., in: M. Luby, J. Rolim, M. Serna (Eds.), Proc. RANDOM'98, Lecture Notes in Computer Science, Vol. 1518, Springer, Berlin, 1998, pp. 260–279.
[16] K. Steinhöfel, A. Albrecht, C.K. Wong, Two simulated annealing-based heuristics for the job shop scheduling problem., European J. Oper. Res. 118 (3) (1999) 524–548.
[17] P.J.M. Van Laarhoven, E.H.L. Aarts, J.K. Lenstra, Job shop scheduling by simulated annealing, Oper. Res. 40 (1) (1992) 113–125.
[18] D.P. Williamson, L.A. Hall, J.A. Hoogeveen, C.A.J. Hurkens, J.K. Lenstra, S.V. Sevast'janov, D.B. Shmoys, Short shop schedules, Oper. Res. 45 (1997) 288–294.